

Continuing Medical Education

"Does ligase chain reaction assay of urine in the diagnosis of *Chlamydia trachomatis* offer significant improvement over existing diagnostic tests?"—a critical appraisal of the evidence

Lucia Grun, Jonathan Sheldon

Introduction

This article sets out a rational, systematic method with which you can assess studies describing new diagnostic tests, and decide whether to incorporate those tests into your clinical practice. Like the first article in this series¹ it uses a current paper as a working example and draws on guides to reading medical literature that appeared recently in the *Journal of the American Medical Association*.^{2,3} The study assessed here looks at urine testing for chlamydial infection using the ligase chain reaction (LCR),⁴ an innovation that could increase the possibility of widespread screening for chlamydial infection outside sexually transmitted disease (STD) clinic settings.

Background

Genitourinary infection with *Chlamydia trachomatis* is the commonest bacterial sexually transmitted disease in the UK.⁵ In women it can have serious long term sequelae, such as pelvic inflammatory disease, chronic pelvic pain and infertility, which have significant personal, social and financial costs.⁶ Because up to 70% of infection in women may be asymptomatic, early detection of lower genital tract infection poses a challenge for clinicians.⁷ Recently there has been much discussion about the possibility of screening women for asymptomatic infection,⁸ especially those thought to be at particular risk of complications: for example, those having a termination of pregnancy.⁹ However, most existing tests are poorly sensitive, that is, fail to identify all of those infected¹⁰⁻¹² and require cervical or urethral samples to be collected.

The ligase chain reaction is a relatively new test that uses gene amplification techniques to

detect chlamydial infection. It has been used successfully on urine samples in men¹³ but its performance in urine samples in women is currently under scrutiny. LCR using urine samples could have two important benefits. Firstly, urine samples are relatively easier to obtain than cervical samples, which would make screening more acceptable to women and widen the range of settings in which testing could occur. Secondly, if LCR is more sensitive and specific than existing tests then fewer infections would be wrongly diagnosed through its use.

The clinical problem

Screening or testing for chlamydial infection outside clinic settings is infrequently performed. This is partly because of the poor sensitivity of tests which are currently available and the need for performing a vaginal examination to collect specimens. The article examined here concludes that LCR is more sensitive than existing tests and could be used for more generalised screening. Using critical appraisal, we test these conclusions, and whether they should influence our clinical practice.

How to approach a paper

When faced with a paper on a new diagnostic test that might potentially change one's clinical practice there are three questions which need to be addressed (Table 1)²:

1. *Are the results of the study valid?* How has the study been carried out, particularly with respect to patient selection and the choice of a "gold standard"? If the study is valid, ask the next two questions.
2. *What are the results of the study?* How sensitive and specific is the new test? Are other measures of performance presented (for example, likelihood ratio), or can these be calculated from the data provided? If not, why not?
3. *Can the results of the study help in the care of my patients?* The acid test of any medical advance is whether it will improve patient care. This is as true for a diagnostic test as a new medical or surgical treatment.

1. Are the results of the study valid?

Primary guides:

Is there an independent blind comparison with a reference standard? Until recently the "gold

Mortimer Market
Centre, Mortimer
Market, off Capper
Street, London
WC1E 6AU, UK
L Grun

Keats Group Practice,
18 Rosslyn Hill,
London NW3, UK
J Sheldon

Address correspondence to:
Lucia Grun, at Mortimer
Market Centre.

Accepted for publication
19 August 1996

Table 1 Evaluating and applying the results of studies of diagnostic tests [1]

Are the results of the study valid?

Primary guides:

- Was there an independent blind comparison with a reference standard?
- Did the patient sample include an appropriate spectrum of patients to whom the diagnostic test will be applied in clinical practice?

Secondary guides:

- Did the results of the test being evaluated influence the decision to perform the reference standard?
- Were the methods for performing the test described in sufficient detail to permit replication?

What were the results?

Are likelihood ratios for the test results presented or data necessary for their calculation provided?

Will the results help me in caring for my patients?

Will the reproducibility of the test result and its interpretation be satisfactory in my setting?

Are the results applicable to my patient?

Will the results change my management?

Will the patients be better off as a result of the test?

standard" for the detection of *Chlamydia trachomatis* has been cell culture of endocervical specimens. There is now evidence that between 5% and 30% of chlamydial infections are in the urethra only and not detected on an endocervical specimen.¹⁴ Furthermore, work with tests based on antibody detection suggests that culture does not detect all cases of infection.¹⁵⁻¹⁷ Because of these limitations the concept of an "expanded gold standard" has been developed. This allows for culture-negative samples to be considered positive if shown to be positive by two or more unrelated tests. Different investigators use slightly different "expanded gold standards," depending on the tests available to them and their own areas of excellence.

In the study under critical appraisal, matched pairs of endocervical and urine specimens were compared. Endocervical samples were tested using culture. First void urine samples were tested using a plasmid based LCR. Staff were blinded to the results of tests when performing the assays. Where the urine LCR was positive and the endocervical culture was negative, these culture-negative endocervical samples were subjected to additional testing in a blinded manner with direct immunofluorescence (DIF) which can directly visualise chlamydial inclusion bodies. If the endocervical specimen was confirmed to be negative on testing with DIF, then LCR positive urine specimens were retested using another LCR assay based on different DNA probes in order to establish whether the LCR result represented a false-positive. When LCR on urine was negative and culture on endocervical specimens was positive, the LCR was considered to be false-negative.

We conclude that the study under appraisal uses a convincing gold standard and has performed its tests in an appropriately blinded fashion.

Does the sample include an appropriate spectrum of those participants to whom the diagnostic test will be applied in clinical practice? A test's usefulness depends on its accuracy in a wide range of clinical settings. A test that only detects *Chlamydia trachomatis* in women with marked symptoms and probable disease would not be useful, for example, in potential screening programmes. Existing tests, such as enzyme immunoassay perform poorly in low prevalence populations, which limits their use as screening tools. If patients either at low risk or high risk of infection are not represented in a study, the full range of test performance will not be apparent. In this study, patients were recruited from four different centres in North America. In only one of these centres was the prevalence of chlamydia over 10%, the remainder having a prevalence between 3.3% and 7%. These levels are consistent with published community prevalence figures in the USA and UK^{18 19} and are therefore likely to be fair representations of the average population.

Secondary guides:

Do the results of the test being evaluated influence decisions to perform reference standard tests? In

some studies, the decision to perform a gold standard test is influenced by the results of the test under scrutiny, particularly where the gold standard test is complex or invasive. For example: performing a pulmonary angiogram to exclude a pulmonary embolism might be considered unethical when the non-invasive test under scrutiny (such as, a ventilation-perfusion scan) is negative. In such studies, the gold standard test is only performed when the probability of a positive result is high, as it might otherwise put patients at unacceptable risks, or be too expensive.

In the study under review, all patients had both endocervical culture and urine LCR performed. The decision to perform the "expanded gold standard" was determined by the presence of a positive urine LCR when the endocervical culture was negative. It ignores the possibility that endocervical culture and urine LCR may both miss true cases of chlamydial infection. This could be the case if a woman had only a urethral infection, which would not be detected on endocervical culture. Some studies have taken urethral and endocervical samples for culture, but this is costly and uncomfortable for patients.²⁰ This does, however, clarify the results in cases of purely urethral infection.

Are the methods for performing the test described in enough detail to permit replication? In the study under review, the methods are described in detail, particularly the LCR testing, which is a novel test with few previous applications described in the literature.

In summary, this study meets the criteria for validity set out above. Once satisfied that the results in a given study are valid, the next step is to look at what they are, and how the new test compares to existing tests and the gold standard.

2. What are the results of the study?

There are a number of different measures of performance of a diagnostic test. Sensitivity and specificity are commonly used, although these do not take into account the population on whom the test is being used. A test that performs well on women attending an STD clinic may perform less well in a different setting where the prevalence of infection is lower.

From a clinical point of view the usefulness of a test depends on how it will improve patient care and at what cost. An ideal test is one that can be relied upon to give an accurate result, whatever the likelihood of infection. However, few tests perform this well. It is therefore useful to have a clear idea of how accurate a test is likely to be when a patient's clinical picture is taken into account.

When we evaluate a patient by taking a history and performing a physical examination, the probability of them having a given disease can be estimated. This probability is called the "pre-test probability." While clinicians rarely make formal estimates of pre-test probability, they nevertheless subconsciously estimate how likely it is that patients have a given disease when taking a history and examining them.

There is also literature quantifying the accuracy of history and examination in determining pre-test probability.²¹

The likelihood ratio (LR) is a numerical tool which allows clinicians to build on their clinical judgment (that is, their estimated pretest probability). The LR is calculated using the same figures as those for calculating sensitivity and specificity. In order to calculate the LR we need to know the proportion of true positives and true negatives identified by the test in question. Using figures in the paper under appraisal, we can construct the 2 × 2 table (shown in table 2).

Looking first at the samples that tested positive by LCR, 149 of the 159 positives, identified by the "expanded gold standard", were correctly identified using LCR. This is a likelihood of 149/159 (= 0.94). Only one sample that was negative according to the gold standard was identified as positive by LCR, giving a likelihood of 1/1777 (= 0.00056). The likelihood ratio of a positive LCR is the ratio of these two figures, in this case 0.94/0.00056 or 1678. In other words, a positive LCR result is 1678 times as likely to occur in a patient with chlamydia infection than in a patient without.

In a similar way we can look at samples that tested negative for chlamydia by LCR to calculate the negative likelihood ratio. This will give us an estimate of how powerful a negative test result is. Using LCR the likelihood of a false negative is 10/159 (= 0.0628). Since only one true negative was identified as positive with LCR, the likelihood of a true negative is 1777/1778 (0.999). Thus the negative likelihood ratio of the test is 0.0628/0.999 (= 0.0629). This means that a negative test result using LCR is 0.0629 times as likely to occur in a patient who has chlamydial infection than in a patient who does not.

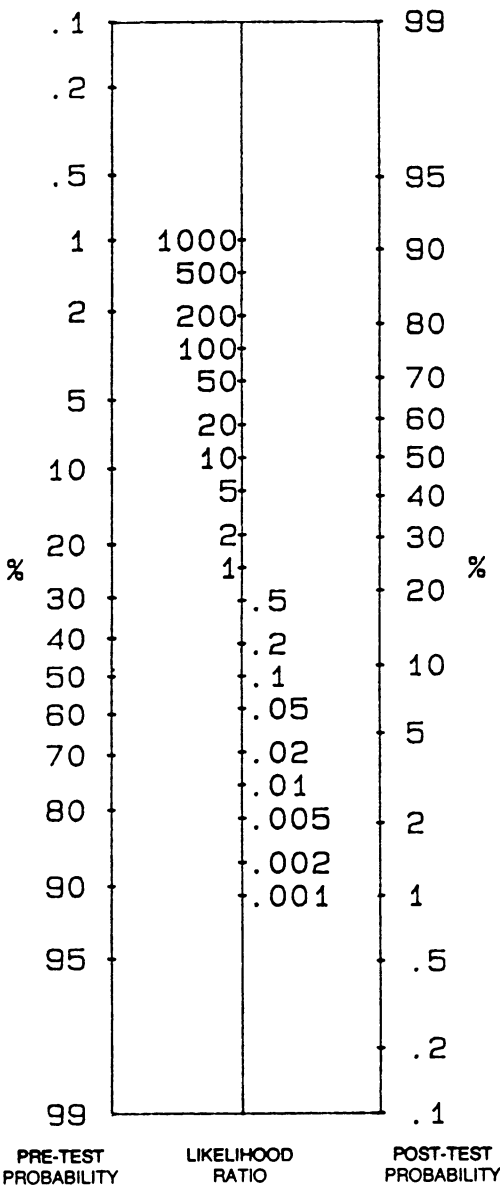
We can use the LR to find out how much the result of a diagnostic test will alter the pre-test probability of a given diagnosis. An LR of 1 means that there will be no change: performing the test will not give any new information nor alter your clinical judgment as to whether the patient has the disease or not. A test with an LR of greater than 1 increases the probability that the patient has the disease over clinical judgment alone and an LR of less than 1 decreases this probability. There is no hard and fast rule about what size LR is clinically useful. However, as a rough guide, LRs of greater than 10 or less than 0.1 can change a pre-test probability significantly. LRs between 1 and 2, or 0.5 and 1, alter the probability to only a very small extent and are unlikely to be important in clinical practice. Intermediate figures can generate small, but possibly important, changes in probability.

Once we know the LR of a test, and whether it is likely to have clinical significance,

Table 2 The 'expanded gold standard'

		Positive	Negative	Total
LCR	Positive	149	1	150
	Negative	10	1777	1787
	Total	159	1778	1937

NOMOGRAM* FOR INTERPRETING
DIAGNOSTIC TEST RESULTS



*Adapted from FAGAN TJ: Nomogram for Bayes's Theorem (C). N Engl J Med 1975; 293: 257.

we need to combine it with the pre-test probability to decide whether the test will alter our clinical management. Combining likelihoods requires cumbersome calculation.³ Fortunately, Fagan²² has devised a simple nomogram (fig) which does all the work. This allows us to combine pre-test probability with the LR of a test to give a post-test probability (that is, the probability of a diagnosis once information from the test result has been incorporated). The first column of the nomogram shows the pre-test probability, the second column gives the LR and the third gives post-test probability. To find out if a given test would alter your clinical judgment, anchor a ruler on the pre-test probability. Then rotate it to pass through the LR for the test in question and read off the post-test probability in the third column.

In the example under appraisal, the LR of LCR is so large that for *any* pre-test probability, however large or small, the likelihood of infec-

tion in the presence of a positive test is extremely high. For example, an asymptomatic patient attending a pre-termination clinic, who has a pre-test probability of infection of 15–20%, would have over 99% probability of infection if her LCR test was positive. Conversely, the small negative LR means that if her LCR test was negative, it is very unlikely that she actually has chlamydial infection: in the example given this would be only 1.5%.

Having established that a diagnostic test gives valid results and what the LR of the test is, we turn to the question of whether the test is useful in clinical practice.

3. Can the results of the study help in the care of my patients?

Will the reproducibility of the test result, and its interpretation, be satisfactory in my setting? The usefulness of any test depends on its reproducibility in different settings. Reproducibility may be poor because of problems with the test itself—for instance sensitivity to temperature or contamination—or because it requires interpretation, which may be subject to personal differences. An article about a diagnostic test should give some guide to how reproducible it is and whether particular problems can be anticipated.

In the study under appraisal, the authors detail how positive results were defined with LCR testing. Furthermore, they specify exactly how they avoided contamination of specimens. In reality, few clinicians will be in a position to order a new test involving new technology until it becomes available locally, at which point these detailed descriptions become important.

Are the results applicable to my patients? It is important for the clinician to know that the test will have the same accuracy in her or his patients as in those studied. If the patients described in a paper are similar to patients seen by the clinician, it is likely that the results will apply. Hence the point made above about the importance of a representative population having been examined in the study.

As already mentioned, the authors of this study have looked at populations with a range of chlamydial prevalence, covering most clinical practice that STD physicians and general practitioners are likely to encounter. We can therefore use its results with confidence, once the technology becomes available.

Will the results change my management? Critical appraisal of this paper suggests that, in experienced hands, LCR testing is an exceptionally promising tool with properties that make its results definitive. If generally available to clinicians at an acceptable cost, it seems likely that urine LCR could become the test of choice for *Chlamydia trachomatis* in men and women. On this evidence, LCR would detect an additional 45 cases of chlamydia compared with culture of endocervical specimens, an increase of 43%. LCR made one false positive diagnosis, which represents a rate of approximately 1 in 200. Clinicians need to decide if this rate of false positives is accept-

able in their clinical practice. LCR is unusual in having such a high LR. More often, clinicians face decisions where a test has less power: while increasing the post-test probability, it will not make the diagnosis certain.

In STD clinics, the poor sensitivity of currently available tests may mean that some women are being falsely reassured that they do not have chlamydia infection. Clinicians should therefore have clear guidelines about what post-test probability of infection is necessary for treatment to be initiated and at what post-test probability they will say an infection is absent. If a result falls between these two levels further investigation is warranted to confirm the diagnosis. Knowing the LR of a diagnostic test allows post-test probability to be determined and decisions to be made in a rational, consistent way. At present in STD clinics it is often appropriate to treat chlamydial infections epidemiologically (that is, through sexual contacts), in part because of the limitations of existing tests. This may not be necessary with a more sensitive and specific test.

When evaluating papers on diagnostic tests, it is unlikely that we will be making choices for individual patients. Where policy decisions are being considered, with significant long-term resource implications, it is essential that such studies are evaluated critically within a structured framework. Results of more than one study need to be considered, and any differences examined carefully and compared to the local conditions.

Conclusion

This paper has provided guidelines for critical appraisal of studies describing diagnostic tests. Using these, you can decide whether studies provide a credible estimate of the value of the test, what the properties of the test are, and whether it is of use to you. Increasingly, articles include the likelihood ratio of the test under scrutiny or at least provide figures that can be used to compute the LR. We have shown how this can be used in clinical practice.

From critical appraisal of the study used as an example here, LCR on urine samples has extremely high positive and negative LRs. It provides clear evidence of the presence or absence of chlamydial infection and, if available to clinicians, it would probably be the test of choice. Although only one paper, the evidence presented here is compelling. Critical appraisal of further studies should influence important decisions about whether the technology needed for LCR testing should be widely introduced and how this could be done.

- 1 Stephenson JM, Williams IG. Is rifabutin prophylaxis against *Mycobacterium avium* complex infection in HIV infection worthwhile? The net impact on patients suggests not. *Genitourin Med* 1996;7:272–6.
- 2 Jaeschke R, Guyatt G, Sackett DL. Users' guides to the medical literature: iii. How to use an article about a diagnostic test: a. Are the results of the study valid? *JAMA* 1994;271:389–91.
- 3 Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the

- medical literature: iii. How to use an article about a diagnostic test: b. What are the results and will they help me in caring for my patients? *JAMA* 1994;271:703-7.
- 4 Lee HF, Chernesky MA, Schachter J, Burczak JD, Andrews WW, Muldoon S, *et al.* Diagnosis of *Chlamydia trachomatis* genitourinary infection in women by ligase chain reaction assay of urine. *Lancet* 1995;345:213-6.
- 5 Governmental Statistical Service. Sexually transmitted diseases, England 1995. New cases seen at NHS genito-urinary medicine clinics. *Department of Health Statistical Bulletin* 1996;July:1-15.
- 6 Berman SM, Campbell CH, Geissman K, Johnson RE, Kennedy EJ Jr, Lossick JG, *et al.* Recommendations for the prevention and management of *Chlamydia trachomatis* infection, 1993. *MMWR* 1993;42:1-37.
- 7 Centers for Disease Control. Chlamydia trachomatis infection: policy guidelines for prevention and control. *MMWR* 1985;34:53-74.
- 8 Scholes D, Stergachis A, Heidrich FE, Andrilla H, Holmes KK, Stamm WE. Prevention of pelvic inflammatory disease by screening for cervical chlamydial infection. *N Engl J Med* 1996;334:1362-6.
- 9 Blackwell AL, Thomas PD, Wareham K, Emery SJ. Health gains from screening for infection of the lower genital tract in women attending for termination of pregnancy. *Lancet* 1993;342:206-10.
- 10 Taylor-Robinson D, Thomas BJ. Laboratory techniques for the diagnosis of chlamydial infection. *Genitourin Med* 1991;67:256-66.
- 11 Thomas BJ, MacLeod EJ, Taylor-Robinson D. Evaluation of sensitivity of 10 diagnostic assays for *Chlamydia trachomatis* by use of a simple laboratory procedure. *J Clin Pathol* 1993;46:408-10.
- 12 Taylor-Robinson D. Tests for infection with *Chlamydia trachomatis*. *Int J STD AIDS* 1996;7:19-26.
- 13 Chernesky M, Lee H, Schachter J. Diagnosis of a *Chlamydia trachomatis* urethral infection in symptomatic and asymptomatic men by testing first void urine in a ligase chain reaction assay. *J Infect Dis* 1994;170:1308-11.
- 14 Hay PE, Thomas BJ, Horner PJ, MacLeod E, Renton AM, Taylor-Robinson D. *Chlamydia trachomatis* in women: the more you look, the more you find. *Genitourin Med* 1994;70:97-100.
- 15 Lefebvre J, Laperiere H, Rousseau H, Marre R. Comparison of three techniques for detection of *Chlamydia trachomatis* in endocervical specimens from asymptomatic women. *J Clin Microbiol* 1988;26:726-31.
- 16 Wu C, Lee M, Yin S, Yang D, Cheng S. Comparison of polymerase chain reaction, monoclonal antibody based enzyme immunoassay and cell culture for detection of *Chlamydia trachomatis* in genital specimens. *Sex Transm Dis* 1992;19:193-7.
- 17 Thejls H, Gnarpe J, Gnarpe H, Larsson P-G, Platz-Christensen J-J, Ostergaard L, *et al.* Expanded gold standard in the diagnosis of *Chlamydia trachomatis* in a low prevalence population: diagnostic efficacy of tissue culture, direct immunofluorescence, enzyme immunoassay, PCR and serology. *Genitourin Med* 1994;70:300-3.
- 18 Oakeshott P, Hay P. General practice update: chlamydia infection in women. *Br J Gen Pract* 1995;45:615-20.
- 19 Hillis S, Black C, Newhall J, Walsh C, Groseclose SL. New opportunities for chlamydia prevention: Applications of science to public health practice. *Sex Transm Dis* 1995;22:197-202.
- 20 Ridgway GL, Mumtaz G, Robinson AJ, Franchini M, Carder C, Burczak JD *et al.* Comparison of the ligase chain reaction with cell culture for the diagnosis of *Chlamydia trachomatis* infection in women. *J Clin Pathol* 1996;49:116-9.
- 21 Panzer RJ, Black ER, Griner PF. *Diagnostic Strategies for Common Medical Problems*. Philadelphia Pa: American College of Physicians 1991.
- 22 Fagan TJ. Nomogram for Bayes's Theorem (C). *N Engl J Med* 1975;293:257.